The mechanism of M.HhaI DNA C5 cytosine methyltransferase enzyme: A quantum mechanics/molecular mechanics approach

Xiaodong Zhang and Thomas C. Bruice*

Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106

Contributed by Thomas C. Bruice, February 27, 2006

The mechanism of DNA cytosine-5-methylation catalyzed by the bacterial M.HhaI enzyme has been considered as a stepwise nucleophilic addition of Cys-81-S$^-$ to cytosine C6 followed by C5 nucleophilic replacement of the methyl of *S*-adenosyl-L-methionine to produce 5-methyl-6-Cys-81-S-5,6-dihydrocytosine. In this study, we show that the reaction is concerted from a series of energy calculations by using the quantum mechanical/molecular mechanical hybrid method. Deprotonation of 5-methyl-6-Cys-81-S-5,6-dihydrocytosine and expulsion of Cys-81-S$^-$ provides the product DNA 5-methylcytosine. A required base catalyst for this deprotonation is not available as a member of the active site structure. A water channel between the active site and bulk water allows entrance of solvent to the active site. Hydroxide at 10^{-7} mole fraction (pH = 7) is shown to be sufficient for the required catalysis. We also show that Glu-119-CO_2H can divert the reaction by protonating cytosine N3 when Cys-81-S$^-$ attacks cytosine, to form the 6-Cys-81-S-3-hydrocytosine. The reactants and 6-Cys-81-S-3-hydrocytosine product are in rapid equilibrium, and this explains the observed hydrogen exchange of cytosine with solvent.

computational

The bacterial enzyme M.HhaI catalyzes the reaction of certain DNA cytosine residues with *S*-adenosyl-L-methionine (AdoMet) to provide a DNA 5-methylcytosine (MC) and *S*-adenosyl-L-homocysteine (AdoHcy) (1). The chemistry of this reaction is relatively well studied. The two-step mechanism for the methylation (Scheme 1) was originally proposed by Santi *et al.* (2, 3) for DNA cytosine methyltransferases. In the stepwise mechanism of Scheme 1, formation of 6-Cys-81-S-cytosine anion is followed by C5 nucleophilic displacement of the methyl group of AdoMet to provide the 5-methyl-6-Cys-81-S-5,6-dihydrocytosine (MCD) (Scheme 1).

Although Scheme 1 was supported by many experiments and has been widely accepted, definitive evidence could not be provided to establish the formation of the transient covalent 6-Cys-81-S-cytosine anion (C5A). The intermediate MCD has been crystallized (4), supporting this general scheme. However, the molecular details of each step are poorly understood.

A plausible refinement to the Santi mechanism proposed by Bhagwat (5) involves stabilization of the thiolate adduct by electron delocalization to O2 of cytosine assisted by the electrostatic interaction of the neighboring arginines (Scheme 2).

Alternatively, Verdine and coworkers (6) proposed that nucleophilic addition of Cys-81-S$^-$ is general acid catalyzed by Glu-119-CO_2H to provide a stable enzyme covalent adduct, 6-Cys-81-S-3-hydrocytosine (CHC) (Scheme 3). In a second step, Glu-119-CO_2^- acts as a general-base catalyst to deprotonate N3 of CHC in concert with methylation of C5 to provide the MCD (Scheme 3).

In this study, we employ the QM/MM (QM = self-consistent-charge density functional tight binding) calculations which show that the addition of Cys-81-S$^-$ to C6 of cytosine is uncatalyzed and concerted with C5 methylation by AdoMet (Scheme 4).

A solid argument is presented for HO$^-$ as the agent for deprotonation converting MCD → MC + Cys-81-S$^-$. The kinetic influences of the enzyme electrostatic interactions in ground and transition states are discussed.

Results and Discussion

Active Site. Fig. 1 shows the hydrogen bonding networks in the ground state active site of E·S determined by energy minimization with the QM/MM Hamiltonian. Hydrogen bonds exist between O2 of cytosine and both Arg-165 (1.57 Å) and Arg-163 (1.90 Å), as well as Wat-340 (1.64 Å). N3 of cytosine is hydrogen bonded to Wat-330 and Glu-119-CO_2H at 2.89 Å and 1.90 Å, respectively. Glu-119-CO_2H is hydrogen bonded with N4 of cytosine and Wat-330 at 2.03 Å and 1.65 Å, respectively. Wat-330 is hydrogen bonded with the NH_3^+ group (AdoMet) at 1.73 Å. Wat-340 is hydrogen bonded with Arg-163 at 1.90 Å. Wat-353 is hydrogen bonded with H5 of cytosine (2.45 Å), Gln-82 (1.65 Å), and Wat-354 (1.68 Å). Wat-354 is also hydrogen bonded with oxygen at the phosphate group of guanine (1.59 Å). Hydrogen bonds exist between Arg-165 and ribose O4′ (1.81 Å), O5′ of the phosphate group in cytosine (2.43 Å and 3.03 Å), and O1P of the phosphate group in cytosine (2.36 Å and 2.90 Å). Thus, the net positive charge of Arg-165 is mainly balanced by the phosphate group of cytosine.

Activation Energies of Different Methyl Transfer Pathways. We calculated the activation energies using QM/MM for addition of Cys-81-S$^-$ at C5 of cytosine to provide the Cys-81-S$^-$ anion adduct without Arg-163 and Arg-165 assistance in Scheme 1, and with Arg-163 and Arg-165 assistance as in Scheme 2. The barrier in both cases is >20.0 kcal/mol for the formation of Cys-81-S$^-$ anion adduct. Thus, the two stepwise mechanisms (Schemes 1 and 2) cannot be involved in the formation of MC.

The reaction of Scheme 3 involves Glu-119-CO_2H protonation of N3 in concert with addition of Cys-81-S$^-$ to C6 of cytosine to provide E·CHC. The QM region depicting the structure of E·CHC is given in Fig. 2.

The formation of E·CHC is associated with a very small activation energy (≈2.2 kcal/mol) by QM/MM. The E·CHC has stability comparable with the reactants and does not undergo methylation in our calculations. This reversible reaction serves as a means of proton exchange (Scheme 5) in cytosine, which is a side reaction during methylation (7).

These results are in excellent agreement with the experimental pre-steady-state kinetic results (7) requiring a reversible side reaction.

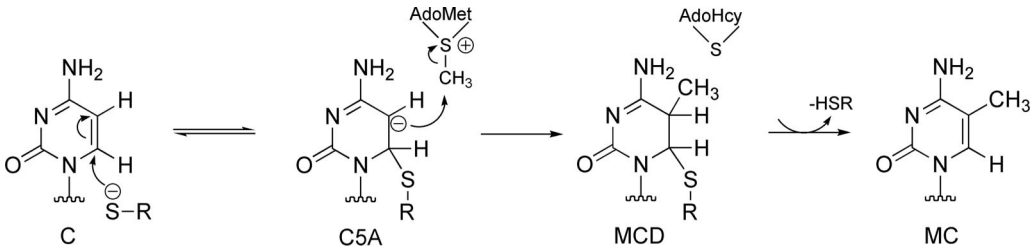
The calculation of potential energy profiles of Schemes 3

Conflict of interest statement: No conflicts declared.

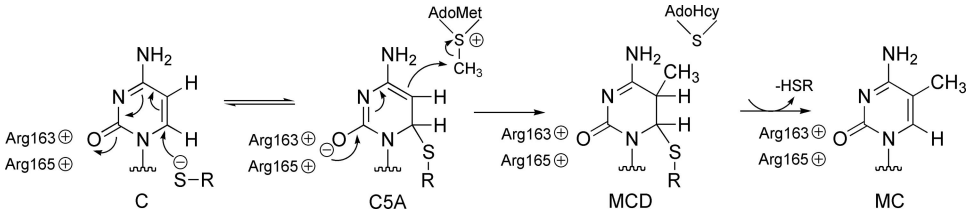
Abbreviations: AdoMet, *S*-adenosyl-L-methionine; CHC, 6-Cys-81-S-3-hydrocytosine; MC, DNA 5-methylcytosine; MCD, 5-methyl-6-Cys-81-S-5,6-dihydrocytosine; MM, molecular mechanics; QM, quantum mechanics; TS, transition state.

*To whom correspondence should be addressed. E-mail: tcbruice@chem.ucsb.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0601587103



Scheme 1.



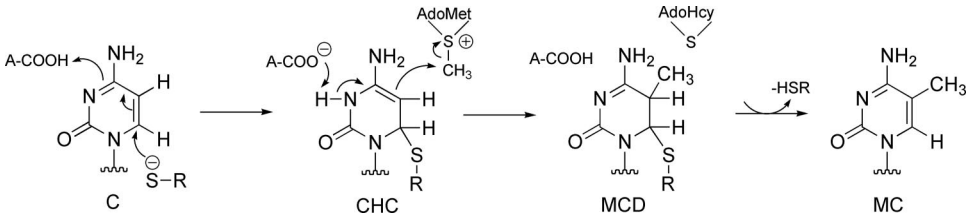
Scheme 2.

and 4 were carried out by using QM/MM (Fig. 3). The QM/MM calculated activation energy for the concerted Cys-81-S^- addition to C6 with methylation of C5 is ≈7.6 kcal/mol. Examination of Fig. 3 indicates that the addition of Cys-81-S^- and methylation reaction is *CHC* ⇔ *C* → *MCD* associated with the activation energy of ≈8.3 kcal/mol, which is similar to that of the experimental enthalpy of the reaction (≈10.3 kcal/mol) (8).

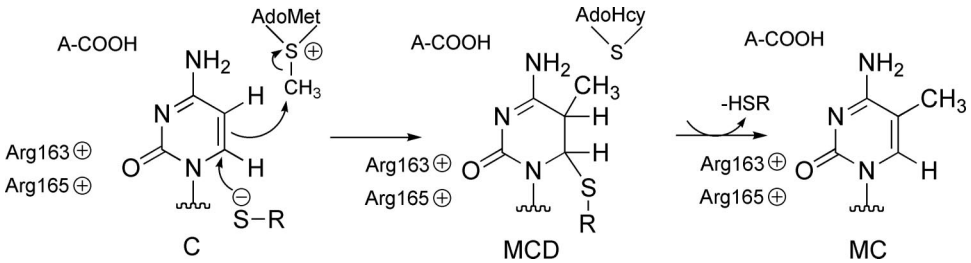
The QM regions depicting the ground state and the transition state (TS-C) for the concerted Cys-81-S^- addition and methylation (Scheme 4) are presented in Figs. 4 and 5, respectively. In the transition state (TS-C) structure (Fig. 5), the S^- (Cys-81) to C6 (cytosine), C5 (cytosine) to C9 (AdoMet), and C9 (AdoMet) to S8 (AdoMet) distances are 2.20 Å, 2.38 Å, and 2.18 Å, respectively.

To understand the effect of electrostatic interaction in stabilizing the TS, the interaction distances between protein residues and the substrate in the ground state were measured and compared with the interaction distances in the TS. The interaction distances in the ground and transition states for the structures in the QM regions are listed in Table 1. The transition state (TS-C) of the concerted reaction occurs at ≈40% reaction. The interactions distances of Glu-119-COOH, Arg-163, and Arg-165 change by 0.02 Å, 0.06 Å, and 0.07 Å, respectively, on going from ground state to transition state for the concerted reaction. Thus, the role of Glu-119-CO_2H is not catalytic but is to create the active ground state conformation. Arg-163 and Arg-165 are also important in creating the reactive conformer and may play a small role in catalysis of bond making and breaking. If the two waters (Wat-330 and Wat-340) in the active site (Fig. 1) are included in the QM regions, there is an insignificant change in the activation energy for the concerted reaction (≈7.75 kcal/mol vs. ≈7.60 kcal/mol), showing that these two waters are not important for catalysis. The separations of electrostatic interactions of Glu-119-CO_2H, Arg-163, and Arg-165 on going from ground state to transition state (see Figs. 8 and 9, which are published as supporting information on the PNAS web site) remain the same. Wat-330 is 0.03 Å closer to Glu-119-COOH in the transition state than in the ground state. There is no difference in the positioning of Wat-340.

The overall reaction of Cys-81-S^- addition and methylation to provide MCD is calculated to be exergonic by as much as 21.2 kcal/mol at the QM/MM level. This is in agreement with the model calculations by Parakyla (9) in the gas phase, who found



Scheme 3.



Scheme 4.

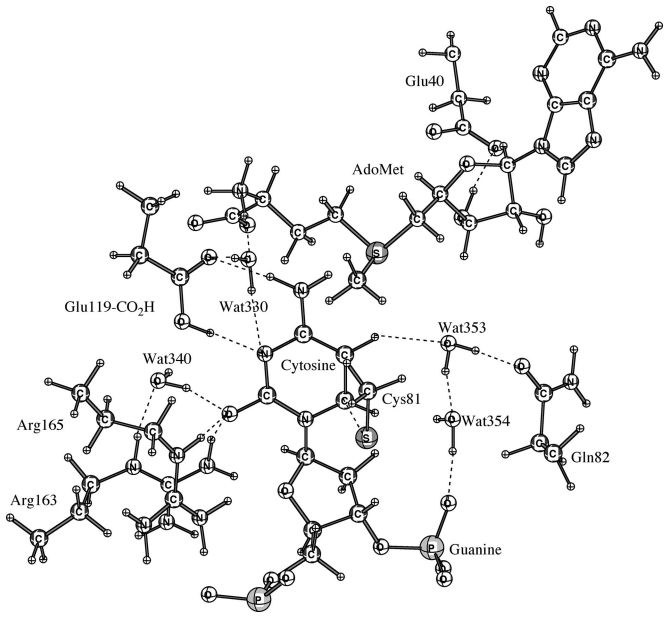


Fig. 1. Diagram of the active-site residues of M.HhaI and crystallographic water molecules around the target cytosine from the minimization structure by using self-consistent-charge density functional tight binding/MM.

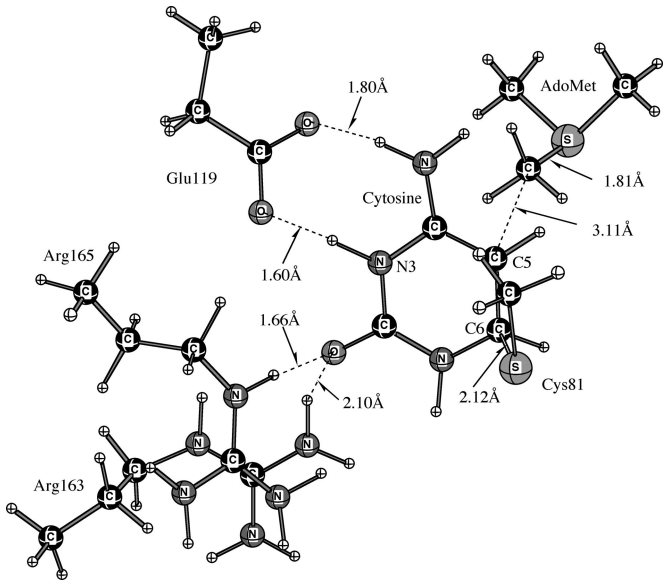


Fig. 2. The QM region depicting the structure of E·CHC (Scheme 3).

the formation of MCD to be exergonic by ≈35.9 kcal/mol, including the solvation energy at B3LYP/6–31+G*. To better understand these energies, we determined the energies for the reaction of Scheme 6 in the gas phase to be exergonic by ≈23.8 kcal/mol at the HF/6–31G* level (Scheme 6).

Thus, changing of C5 and C6 in reactant II from an sp^2 hybridization plane into sp^3 hybridized product III accounts for the difference in stability of cytosine and MCD.

Deprotonation of C5 and elimination of Cys-81-S is the last step in M.HhaI methylation reaction. There is no suitable base near the active site in the crystal structure. Santi *et al.* (2) proposed that H5 could be exchanged with solvent. Our previous molecular dynamics simulations (10) observed that a solvent water channel (X = H_2O; Fig. 6) leads to the active site so that H_2O, H_3O^+, and HO^- have access.

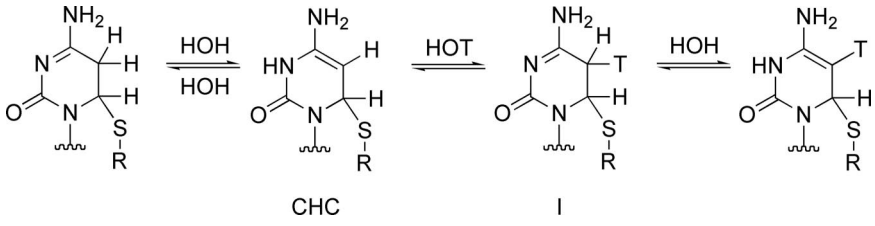
For the study of the proton abstraction and Cys-81-S^- expulsion, when H_2O is assumed to be the base (X in Fig. 6) that removes the proton from C5, we find the reaction energies to be as much as ≈26.5 kcal/mol by means of QM/MM. Thus, H_2O cannot be accepted as the base for deprotonation. A base stronger than water is required to remove this proton. By QM/MM calculation the activation energy with X = HO^- (Fig. 6) at the active site is ≈2.0 kcal/mol. Because HO^- at neutrality is present at the fraction of ≈10^{-7}, the activation energy would be ≈12.0 kcal/mol. This provides a reasonable reaction energy profile as shown in Fig. 3. At pH values >7.0, the barrier will be <12 kcal/mol.

Perturbation Analysis of Electrostatic Effects for Methyl Transfer Process. It is possible that residues that are not directly in contact with the substrate or AdoMet are involved in electrostatic stabilization of the TS. Thus, we analyzed the contribution of electrostatic effect to catalysis by measuring the difference ($\Delta\Delta E^*$) in the calculated activation energies for a residue with its normal partial charge and with no charge (perturbation analysis). The results at the level of QM/MM are shown in Fig. 7.

The perturbation analysis of Fig. 7 for the concerted reaction (Scheme 4) indicates that the dominant and unfavorable effects of Asp-16 (4.30 kcal/mol), Asp-42 (3.53 kcal/mol), and Wat-340 (2.60 kcal/mol) increase the activation energy compared with that in solution or in the gas phase model. These residues stabilize the ground state of E·S. The favorable contributions of Glu-40 (−7.58 kcal/mol), Arg-25 (−1.90 kcal/mol), Asn-304 (−2.22 kcal/mol), and Lys-300 (−2.05 kcal/mol) decrease the activation energy. These residues stabilize the transition state.

Kinetic Effects of Electrostatic Interactions of Residues at Long Range. Reich and coworkers (8) have mutated long range residues and determined the effect of the various mutations on the rate of methyl transfer (Table 2, last two columns). The calculated rate constants of these mutations for the concerted mechanism from the perturbation analysis (Fig. 7) are listed in Table 2.

Examination of Table 2 reveals that Arg-106 and Lys-111 have −0.33 and −0.11 kcal/mol, respectively, favorable contributions to the activation energy in the concerted mechanism. Their corresponding calculated rate constants are 0.08 and 0.12 s^{-1}, respectively. Asp-71, Asp-73, Met-168, and Val-282 have 0.05,



Scheme 5.

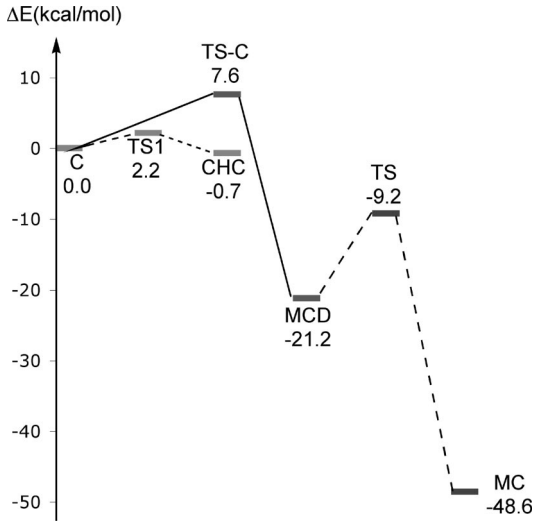


Fig. 3. The schematic effective energy surfaces (in kcal/mol) for the concerted path (solid line), rapid equilibrium (stepwise path) (dotted line), and deprotonation with elimination of Cys-81-S^- (dashed line) at the level of self-consistent-charge density functional tight binding/MM.

0.05, 0.03, and 0.01 kcal/mol unfavorable contribution to the activation energy of the concerted mechanism, with calculated rate constants of 0.15, 0.15, 0.15, and 0.14 s^{-1}. A plot of the calculated rate constant vs. the experimental (8) values provided a linear plot of slope 1.6 ($R^2 = 0.79$). Thus, the calculated and experimental (8) rate constants are in relative agreement.

Conclusions

The mechanism of M.HhaI catalysis of methylation of selected DNA cytosine was known to involve Cys-81-S^- nucleophilic addition to cytosine C6 with methylation at C5 by AdoMet to provide MCD, followed by deprotonation and elimination of Cys-81-S^- to provide MC. The methylation reaction has been considered to be stepwise with or without Glu-119-CO_2H/Glu-119-CO_2^- general acid/general base catalysis and/or electrostatic stabilization of the intermediate by Arg-163 and

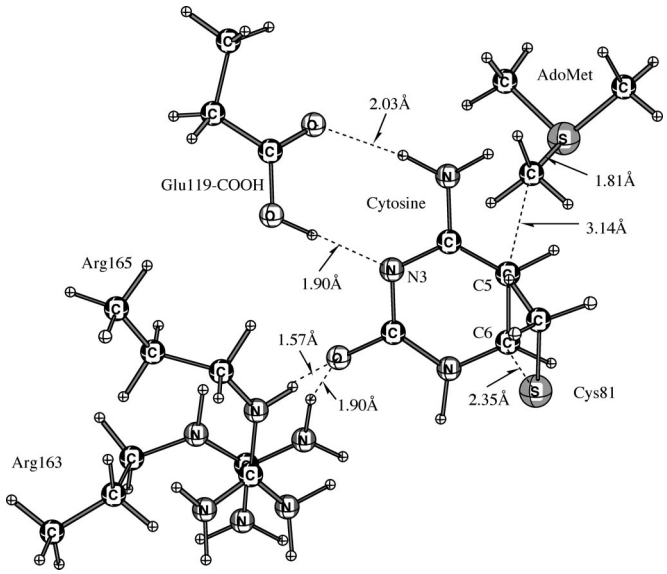


Fig. 4. The QM region depicting the ground state.

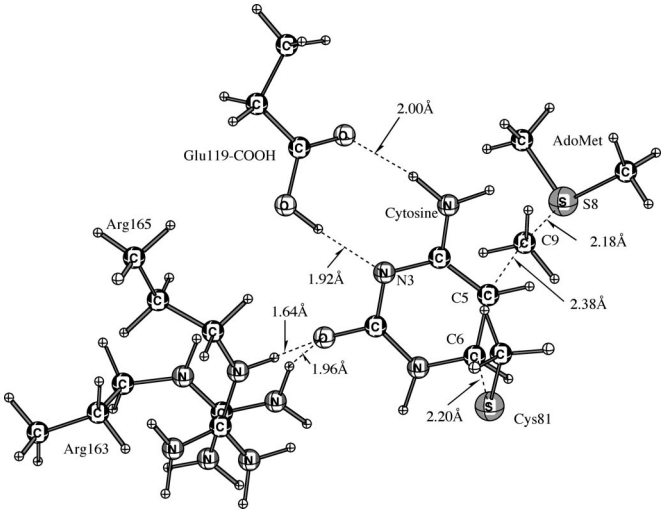


Fig. 5. The QM region depicting the transition state (TS-C) for the concerted mechanism (Scheme 4).

Arg-165 (Schemes 1–3). The base species for deprotonation of MCD has escaped identification.

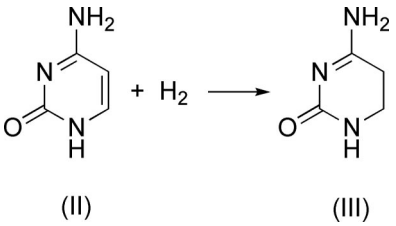
The reactions of interest at the M.HhaI active site in water solvent have been investigated by the QM/MM method with QM = self-consistent-charge density functional tight binding. The activation energies for the concerted Cys-81-S^- nucleophilic addition with AdoMet methylation to provide MCD are associated with a calculated activation energy of ≈8.3 kcal/mol, which is similar to that of the experimental enthalpy (≈10.3 kcal/mol). Comparison of electrostatic bond lengths in the ground state and the transition state establishes a lack of catalysis by Glu-119-CO_2H and a minor assistance by Arg-163 and Arg-165. These electrostatic interactions, however, are of major importance in creating the reactive ground state conformer (NAC). There has been, for some time, the question of the nature of the base responsible for deprotonation of MCD → Cys81SH + 5-methylcytosine. There is no appropriate base at the active site. A water channel (described in ref. 10) leads to the active site. Calculations provide the activation energy (≈2.0 kcal/mol) for the deprotonation by a neighboring HO^-. At pH 7, this barrier would be ≈12.0 kcal/mol, which is quite reasonable (Fig. 3). The calculated rates from the perturbation analysis for the long-range residues are in good agreement with the experimental values.

In a side reaction at the active site of M.HhaI, Glu-119-CO_2H protonates the N3 of cytosine in concert with Cys-81-S^- addition to C6. The product CHC is in rapid equilibrium with the reactants and does not undergo methylation. It is suggested that proton exchange from CHC and water solvent is the

Table 1. The important distances (in Å) around the target cytosine at the ground state, the transition state of the concerted mechanism, and the difference

	Ground state	Transition state	Difference
OE1(Glu-119) to H41(cytosine)	2.03	2.00	−0.03
HE2(Glu-119) to N3(cytosine)	1.90	1.92	0.02
HH21(Arg-163) to O2(cytosine)	1.90	1.96	0.06
HE(Arg-165) to O2(cytosine)	1.57	1.64	0.07
SG(Cys-81) to C6(cytosine)	2.35	2.20	−0.15
C9(AdoMet) to C5(cytosine)	3.14	2.38	−0.76
C9(AdoMet) to S8(AdoMet)	1.81	2.18	0.37

BIOCHEMISTRY



Scheme 6.

mechanism for the known hydrogen exchange into cytosine during the time of the methylation reaction.

Computational Method

The QM/MM [QM = self-consistent-charge density functional tight binding (11, 12)] approach was implemented in CHARMM (Version 31b1) (13). The starting structure for the ground state was built from the 2.05-Å-resolution x-ray structure [PDB entry 6MHT (14)] of the tertiary complex of enzyme, DNA, and AdoMet, by replacing the active site 4′-thio-2′-deoxycytidine with cytosine. The initial structure was solvated in a box of TIP3P water. The stochastic boundary (15) of radius 25 Å was centered at the cofactor AdoMet. Those atoms beyond 25 Å were deleted. The final model included the AdoMet cofactor with 50 atoms, 4,053 protein atoms, 665 DNA atoms, 113 x-ray crystal water molecules, and 741 TIP3P water molecules. A Poisson–Boltzmann charge-scaling scheme (16) was used to include the correction of long-range electrostatic interactions in the simulation. Poisson–Boltzmann calculations determined a set of scaling factors, which reduced the partial charges of charged residues in the QM/MM electrostatic potential calculations so as to avoid artifactual structural change.

The QM region in the QM/MM calculations included $-CH_2$-S^+(Me)-CH_2- of AdoMet cofactor, $-CH_2S^-$ of Cys-81, and the pyrimidine ring of cytosine in DNA substrate for Scheme 1. The side chains Arg-163, Arg-165, and Glu-119-CO_2H are included in the QM region for the reaction in Schemes 2–4. The link atoms (i.e., the hydrogen atoms) were introduced to saturate the valence of the QM boundary atoms. The prepared structure for the ground state was then equilibrated by running a molecular dynamics simulation with QM/MM Hamiltonian for 120 ps with 1-fs time step. The ground state structure was obtained by energy-minimizing the final structure from this molecular dynamics simulation by using the adopted basis Newton–Raphson method [the tolerated gradient was set to 0.01 kcal/(mol·Å)]. The transition state was obtained by using the adiabatic mapping method and confirmed by frequency analysis, which provided only one imaginary frequency. The product was obtained by adopted basis Newton–Raphson minimization of the final structure on the potential surface. To analyze the effects of the side chain in the individual residues on the reaction barrier, the perturbation treatment similar to that used in triosephosphate isomerase (17) was performed.

We studied the side exchange reaction pathway by analyzing the energy surface with two-dimensional reaction coordinates; one is the distance between S of Cys-81 and C6 of cytosine, and the other is the difference between C5 (cytosine) to H5 and H5 to O (Wat-353). The QM regions included Wat-353, Wat-354, X (X = H_2O or HO^- in Fig. 6), $-CH_2S^-$ of Cys-81, $-CH_2$-S^+(Me)-CH_2- of cofactor (AdoMet), side chains of Arg-163 and Arg-165, and cytosine, as well as the phosphate group of the neighboring guanine.

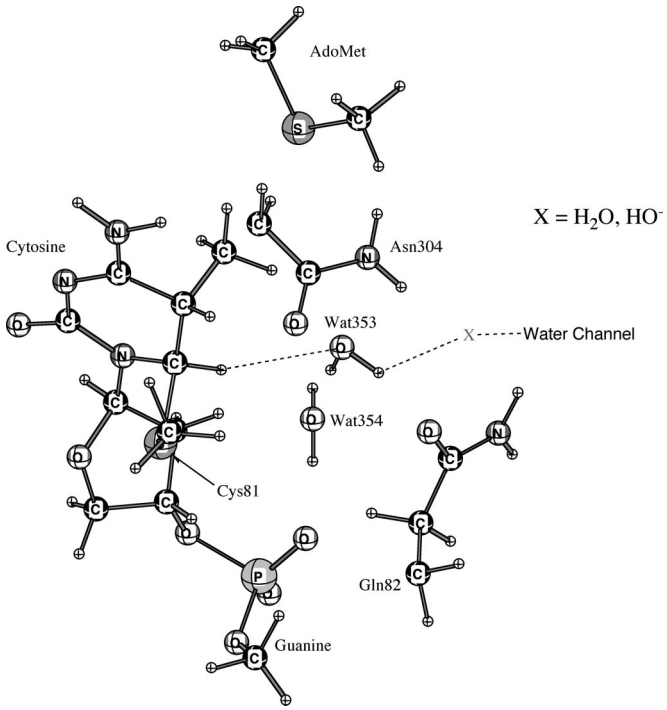


Fig. 6. Arrangement of water channel in close proximity to H5 (cytosine).

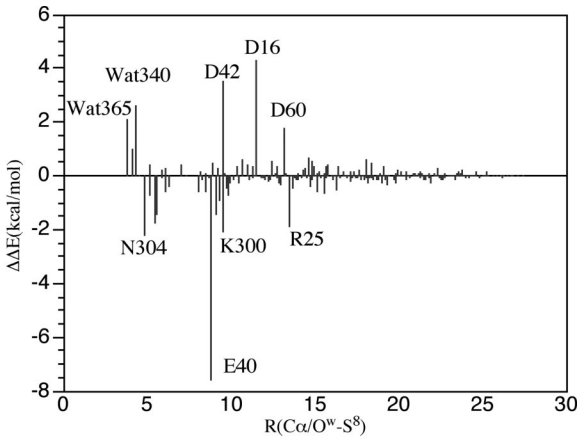


Fig. 7. Perturbation analysis of the electrostatic interactions of residues and crystal lattice water for the concerted mechanism as in triosephosphate isomerase (17).

Table 2. The calculated rate constants from perturbation analysis of the concerted mechanism and experimental mutations for long-range residues

	Computational results*			Experimental values†	
Distance: R(Cα-S⁸), Å	Residues	ΔΔ*E*	*k*, s^{-1}	Mutations	$k_{methyltransfer}$, s^{-1}
23.31	Asp-71	0.05	0.15	Asp71Ala	0.20
21.90	Asp-73	0.05	0.15	Asp73Ala	0.15
19.91	Arg-106	−0.33	0.08	Arg106Ala	0.07
22.76	Lys-111	−0.11	0.12	Lys111Ala	0.13
17.57	Val-282	0.01	0.14	Val282Ala	0.20

*The rates are calculated by using the equation $k = k_{wt}$*exp($\Delta\Delta E/RT$), where $k_{wt} = 0.14\ s^{-1}$.

†The values are taken form experimental values for their mutations (8).

We thank Mr. Joe Toporowski for critically reading the manuscript and Mr. Istvan Szabo for kind help. This work was supported by National Institutes of Health Grant 5R37DK09171. Some of the calculations were performed at the National Center for Supercomputing Applications (University of Illinois at Urbana–Champaign).

1. Kumar, S. H., Horton, J. R., Jones, G. D., Walker, R. T., Roberts, R. J. & Cheng, X. (1997) *Nucleic Acids Res.* **25,** 2773–2783.
2. Wu, J. C. & Santi, D. (1987) *J. Biol. Chem.* **262,** 4778–4786.
3. Santi, D. V., Garett, C. E. & Barr, P. J. (1983) *Cell* **33,** 9–10.
4. O'Gara, M., Klimasauakas, S., Roberts, R. J. & Cheng, X. (1996) *J. Biol. Chem.* **261,** 634–645.
5. Gabbara, S., Sheluho, D. & Bhagwat, A. S. (1995) *Biochemistry* **34,** 8914–8923.
6. Chen, L., MacMillan, A. M. & Versinde, G. L. (1993) *J. Am. Chem. Soc.* **115,** 5318–5319.
7. Svedruzic, Z. M. & Reich, N. O. (2004) *Biochemistry* **43,** 11460–11473.
8. Sharma, V., Youngblood, B. & Reich, N. (2005) *J. Biomol. Struct. Dyn.* **22,** 533–543.
9. Perakyla, M. (1998) *J. Am. Chem. Soc.* **120,** 12895–12902.
10. Lau, E. Y. & Bruice, T. C. (1999) *J. Mol. Biol.* **293,** 9–18.
11. Cui, Q., Elstner, M., Kaxiras, E., Frauesheim, T. & Karplus, M. (2001) *J. Phys. Chem. B* **105,** 569–585.
12. Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Fraucnhcim, T., Suhai, S. & Seifert, G. (1998) *Phys. Rev. B* **58,** 7260–7268.
13. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4,** 187–217.
14. O'Gara, M., Roberts, R. J. & Cheng, X. (1996) *J. Mol. Biol.* **263,** 597–606.
15. Brooks, C. L. & Karplus, M. (1989) *J. Mol. Biol.* **208,** 159–181.
16. Simonson, T., Archontis, G. & Karplus, M. (1997) *J. Phys. Chem. B* **101,** 8349–8362.
17. Cui, Q. & Karplus, M. (2001) *J. Am. Chem. Soc.* **123,** 2284–2290.